2020 2nd International Conference on Information Technology and Computer Application (ITCA)

Deep Learning in Image Classification: A Survey Report

Zhengyu He*
School of Information Science and Technology
Fudan University
Shanghai, China
*17307130174@fudan.edu.cn

***Abstract*—Recently, deep learning is emerging as a powerful tool and has become a leading machine learning tool in computer vision and image analysis. In this survey paper, we provide a snapshot of this fast-growing field, image classification, specifically. We briefly introduce several popular neutral networks and summarize their applications in image classification. In addition, we also discuss the challenge of deep learning in image classification.**

Keywords- Deep learning; Neural network; Image classification

I. INTRODUCTION

Machine learning methods have been inclusively applied to many fields. Achievements in machine learning and artificial intelligence research have brought in new perspectives of conducting neural computation [1]. Compared with non-learning-based techniques that might not translate domain knowledge into rules, machine learning gains its own knowledge from data representations [2]. Nonetheless, traditional machine learning methods generally do not just deal with raw data. They to a great extent rely on the data representations and need comprehensive expertise and subtle layout.

Deep learning techniques are representation-learning methods with compound levels of representation.[3] These algorithm acquire composing rather simple but non-linear modules, and each of them transforms the representation at a single level into a representation at a higher, more essential level [3]. As a result, quite sophisticated functions can be learned through a combination of these transformations. For image classification tasks, the higher-level representation enlarges aspects of data input that are important to distinguish and suppress irrelevant changes. In addition, during the process of image classification, one single image go through several transformations then be sent to the final output. Generally, an image is composed in the form of an array of pixel values. These data that are assigned, contain information of original images. In the first layer, the learned features of representation characteristically represent the presence or absence of edges at particular orientations and locations. In the second layer, it characteristically detects motifs by identifying special arrangements of edges, despite the fact that there are minor variations in the edge. And in the third layer, deeper features could gather into larger combinations which correspond to sections of recognized objects. Furthermore, subsequent layers would detect the same objects. They are combinations of the mentioned recognized objects. The key fact of deep learning algorithm is that all the learned features derived from these various layers are not artificial. That is to say, they are not designed by a human. They are intrinsic features of an array of data, only general-purpose learning procedure could attain them and utilize them to finish tasks [3].

II. NEURAL NETWORKS

The key insight to understand deep learning is to understand neural networks. Artificial Neural Networks (ANN), have been made to contribute in resolving problems of different realms in the recent several decades. Generally speaking, ANN can be depicted as a numerical model of a certain structure, comprising of some of the single processing components, constructed between inter-connected layers [4, 5]. Every unique layer comprises hidden neurons that are used for transforming the input data and figuring out the outputs to the next associated neurons. Deep learning methods have successfully demonstrated their advantages in many applications, where the input values are characterized by high dimensionality, huge quantities and highly-structured. Consequently, deep learning methods are widely applied in image classification and have good performance. Because of the framework of the image that comprises millions of pixels that can be clearly aligned into well-defined objects, deep learning tools are quite useful in the field of image classification [6]. In this section, we will introduce three most popular neural networks.

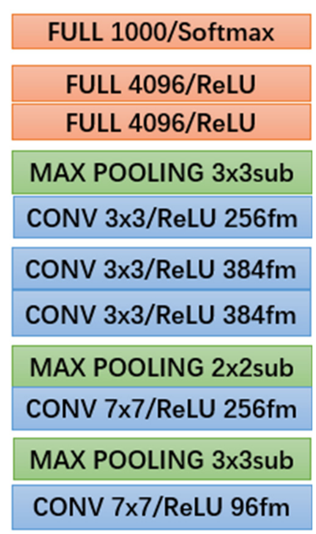

Fig 1. A typical CNN structure.

978-0-7381-1141-4/20/$31.00 ©2020 IEEE
DOI 10.1109/ITCA52113.2020.00043

A. Convolutional Neural Networks (CNN)

One of the most effective models for deep learning is the Convolutional Neural Network (CNN) [3]. The CNN comprises two unique types of layers, pooling layers and convolution. Layers of CNN contain well-designed filters to handle with input data. They convolve the range of input values, and finally get smaller range of them. And then, CNN can detect essential or specific features within the range we acquired before. The CNN generally consists of the input layer, convolution layer and Rectified Linear Unit. Rectified Linear Unit (ReLU) is mathematically expressed as Max(0, x), Max pooling and the final output layer [3], as shown in Fig.2.

Convolution layer can produce a matrix of smaller dimension than input matrix, and max pooling can transmit the maximum value from amongst a rather small batch of data of the input matrix to the output. The output layer is fully-connected, and based on the activation function.

Total input to the j-th feature map of layer l at position (x,y):

$$v_j^{(l)}(x,y) = \sum_{i=1}^{I} \sum_{u,v=0}^{F-1} k_{ji}^{(l)}(u,v) \cdot O_i^{(l-1)}(x-u, y-v) + b_j^{(l)} \tag{1}$$

Convolution layer output:

$$O_i^{(l-1)}(x,y) = f\left(v_j^{(l)}(x,y)\right) \tag{2}$$

Pooling layer output:

$$O_i^{(l+1)}(x,y) = \max_{u,v=0,\ldots,G-1} O_i^{(l)}(x \cdot s+u, y \cdot s+v) \tag{3}$$

Where $O_i^{(l-1)}(i = 1,\ldots,l))$ represents featuremaps on the l+1 layer; where $k_{ji}^{(l)}(u,v)$: indicates trainable convolution kernel; $b_j^{(l)}$ means trainable bias; G is pooling size; and S denotes stride (spacing between adjacent pooling windows).

Convolutional neural networks have achieved great performance in image classification targets [7]. Given input images and corresponding labels, CNN is learned to produce hierarchical representations of data. CNN can impose local connectivity on the raw input value. 1D input data includes signal, 2D input data includes image and voice cepstrum and 3D input data includes tomography and video. The table shows how CNN evolves in recent years.

TABLE I. LIST OF DIFFERENT MODELS

Year	Conference	Model
2012	NIPS	AlexNet
2014	ICLR	NIN
2015	ICLR	VGGnet
2015	CVPR	GoogleNet
2015	ICML	BN
2016	CVPR	InceptionV23
2016	CVPR	ResNet
2016	ECCV	ResNet1001
2016	BMVC	WRNs
2017	CVPR	ShuffleNet
2017	AAAI	Inception V4
2017	CVPR	DenseNet
2017	CVPR	SENet
2017	CVPR	DRN

The list is growing. Till now, CNN is still the state-of-the-art method in many realms.

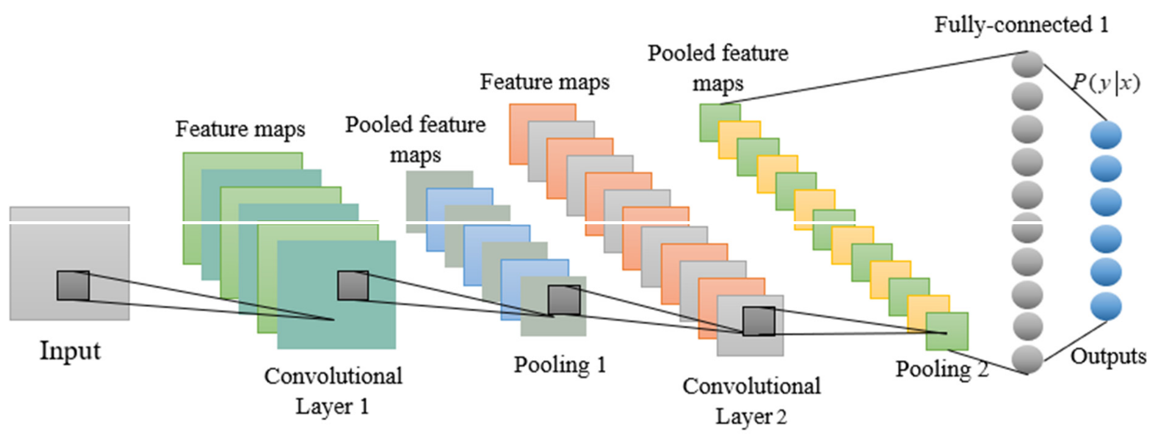

Fig 2. Standard Convolutional Neural Network Architecture generally consists of an input layer, convolution layer, Max pooling and the final fully-connected neural network layer which gives the output based on the activation function.

B. Recurrent Neural Networks

While CNNs are a natural option of tools when the input data are composed of clear spatial structure, especially for pixels, the Recurrent Neural Networks (RNNs) are an appropriate option when data is sequentially aligned (such as natural language or time series) [8]. While simple one-dimensional arrays can be sent to a CNN model, the finally extracted features are quite shallow [8]. Compared with the convolutional neural network which has outstanding performance with the realm of image classification, RNN has demonstrated great efficiency for sequential ordered data in prediction targets and sequence labeling. Two most famous application of RNN is Apple's Siri and Google Voice.

In most traditional neural networks, all the inputs and outputs are usually independent of each other. Apart from other neural networks, RNN contains the history data of input value within the inside neural network. Thus, RNN output has inner connection with past input. RNN imposes the same operations on each element so it is recurrent. An intrinsic feature of RNN is that it has memories which means RNN captures former information. Theoretically RNNs can make use of information in arbitrarily long sequences, but in practice they are limited to looking back only a few steps [9]. This is what a typical RNN structure looks like:

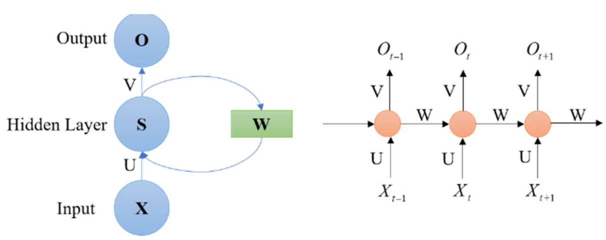

Fig 3. Framework of A Recurrent Neural Networks

Input Layer: X_t Output Layer: O_t Hidden Layer: S_t

Parameter Matrices and Vectors: *U, V, W*

Activation Function of Output Layer: $\boldsymbol{\sigma}_y$

Activation Function of Hidden Layer: $\boldsymbol{\sigma}_h$

The three parameters (*U, V, W,* above) are shared over the whole procession because of numerous repetitive works, corresponding with the work recurrent. As a result, less parameters are needed when we conduct RNN model to learn.

C. Graph Neural Networks

The first appearance of brand-new concept Graph Neural Network is in [10]. This realm set up a new neural network which operate the data represented in graph domains [11]. Unlike other typical neural networks, which can only deal with the data represented in Euclidean space, the GNN can handle inclusive data generated from non-Euclidean domains [12]. These non-Euclidean-domain represented graphs, which have interdependency and relevance inside, can be sent to a GNN model and the related tasks are never unfeasible.

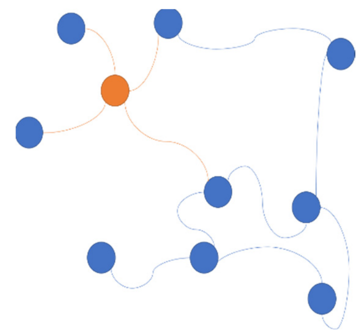

Fig 4. Graph Convolution

The novel method can harness the average value of node features of the orange node along with its neighbors to acquire a hidden representation of the orange node [12]. In CNN or RNN, data are well-ordered and have the same size. Nonetheless, in a GNN model the neighbors of a node is ataxic. In a graph-like data set, node is logically defined by its neighbors (the related nodes) and its intrinsic features.

GNN is emerging as a major tool to resolve some problems such as image classification, and several commonly-used GNN training methods are shown in Fig. 5.

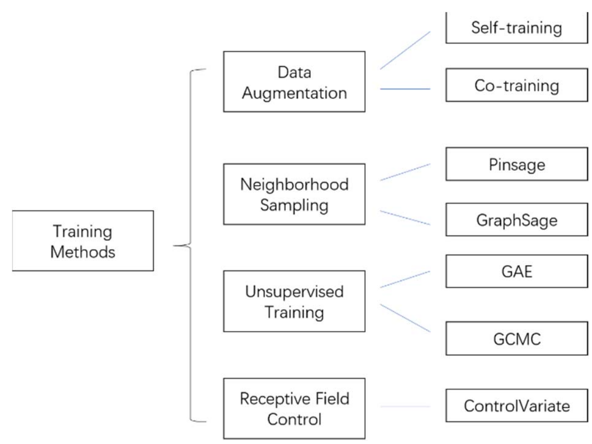

Fig 5. GNN training method

III. IMAGE CLASSIFICATION

Image classification is a compound and comprehensive task that many factors should be taken into account during the process [13]. Recently, more and more new and useful image classification algorithms and techniques are emerging and researchers can evaluate them in terms of classification accuracy and time efficiency. A key insight of a certain image classification target is to selecting a suitable procedure due to the fact that there is not a perfect approach. Different method may have different performance in various tasks. To verify the most optimized procedure towards a specific task, researchers should consider the data type, data size and expected result. In general, a classification system is designed based on spatial resolution of selected remotely sensed data, the user's need, compatibility with previous work, image-processing and classification algorithms available, and time constraints [13].

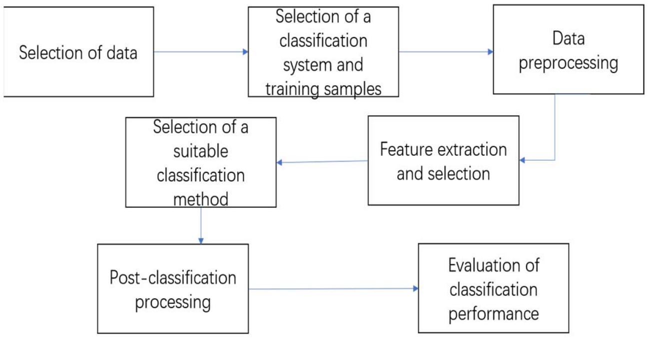

Fig 6. General process of Image Classification

IV. CHALLENGES

Despite the great achievements in image classification, deep learning has following challenges[14]: 1) Number of inputs to be considered and finding non-contributing columns; 2) Number of neurons in each hidden layer; 3) Number of hidden layers; 4) Activation Functions; 5) Optimization Algorithms; 6) Decay function; and 7) Number of epochs and batch size during processing

Deep learning field need continuous progress of these unresolved and ambiguous areas. They bring inevitable challenges both theoretically and experimentally when any deep learning methods are conducted [14].

V. CONCLUSION

Deep learning is a powerful tool for image classification. Different neural networks, such as CNN, RNN, and GNN, play different roles in image classification tasks. To achieve a certain target, it is important for the researcher to select a appropriate deep learning method. However, there are still several challenges in deep learning, and to overcome them can be a major task in the future.

REFERENCES

[1] Andrew Saxe, Stephanie Nelli and Christopher Summerfield, "If deep learning is the answer, then what is the question?", Nature Reviews Neuroscience 22, pp.55-67, 2020.

[2] Fuyong Xing, Yuanpu Xie, Hai Su, Fujun Liu and Lin Yang, "Deep Learning in Microscopy Image Analysis: A Survey", *IEEE Transactions On Neural Networks And Learning Systems*, vol.29, no.10, pp.4550-4568, 2018.

[3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning", Nature 521, pp.436-444,2015.

[4] S.A. Khaparde, P.B. Kale and S.H. Agarwal, "Application of artificial neural network in protective relaying of transmission lines", Proceedings of the First International Forum on Applications of Neural Networks to Power Systems, 1991.

[5] Luciano Caroprese, Pierangelo Veltri, Eugenio Vocaturo and Ester Zumpano, "Deep Learning Techniques for Electronic Health Record Analysis", International Conference on Information, Intelligence, Systems and Applications (IISA), 2018.

[6] Mona Mahrous Mohammed, M. B. Abdelhalim and Amr Badr, "An Optimized PCNN For Image Classification", 10th International Computer Engineering Conference (ICENCO), 2014.

[7] Xiangrong Zhang, Yujia Sun, Kai Jiang, Chen Li, Lichen Jiao and Huiyu Zhou, "Spatial Sequential Recurrent Neural Network for Hyperspectral Image Classification", IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, Vol.11, pp.4141-4155, 2018.

[8] Benjamin Shickel, Patrick James Tighe, Azra Bihorac, and Parisa Rashidi, "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis", IEEE Journal Of Biomedical And Health Informatics, vol.22, pp.1589-1604, 2018.

[9] Jianqiong Xiao and Zhiyong Zhou, "Research Progress of RNN Language Model", IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2020.

[10] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," IEEE TNN 2009, vol. 20, no. 1, pp. 61–80, 2009.

[11] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun, "Graph Neural Networks: A Review of Methods and Applications", arXiv:1812.08434, 2018

[12] Zonghan wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang and Philip S. Yu, "A Comprehensive Survey on Graph Neural Networks", IEEE Transactions on Neural Networks and Learning systems, 2020.

[13] D.Lu, and Q.Weng, "A survey of image classification methods and techniques for improving classification performance," International Journal of Remote Sensing, vol. 28, pp. 823–870, 2007.

[14] Ochin Sharma, "Deep Challenges Associated with Deep Learning," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019.